

早稲田大学大学院 理工学研究科

博 士 論 文 概 要

論 文 題 目

ベイズ決定理論による質問からの学習問題
に関する研究

A Study on Learning from Queries based on
Bayesian Decision Theory

申 請 者

浮田	善文
Yoshifumi	Ukita

--

2007 年 7 月

今日の情報化社会において、高度な情報処理が可能なコンピュータは必要不可欠なものとなっている。人間は与えられた例や事実から法則や一般的な規則を導き出すことができるが、同様の学習能力をコンピュータに持たせることで更なる高度な情報処理が可能となる。コンピュータによる学習に関する研究は、機械学習と呼ばれ、これまで人工知能の研究分野で盛んに研究が行われている。

機械学習のモデルは学習対象に関する情報の与えられ方により、例からの学習と質問からの学習の2種類に大別される。例からの学習とはランダムサンプリングなどにより受動的に得られる情報から学習するモデルである。一方、質問からの学習は1988年に Angluin により提案され、学習機械が質問を行うことにより能動的に情報を得ることを許すモデルである。情報収集にかかるコストを考慮すると、必要な情報を効率よく集めることが可能な質問からの学習が特に重要であり、近年盛んに研究されている。本研究でも質問からの学習を対象とする。ここで、一般に質問回数の増加につれて推定の精度は高くなり、学習効率と学習精度にはトレードオフの関係が存在する。質問からの学習ではこのトレードオフ関係が重要となる。

学習対象を考えた場合、代表的な学習対象の1つとしてブール関数があげられる。ここで考える質問からのブール関数の学習問題とは、未知である真のブール関数に対し、学習機械は自ら選択した入力について所属性質問を行うことで入出力対を得ることができ、いくつかの入出力対を得た後で真のブール関数を推定する問題である。このとき、質問からの学習アルゴリズムは、

(1)次に質問する入力の決定方法

(2)さらに質問を続けるか、それとも質問を終了するかの決定方法

(3)質問後の仮説の決定方法

により構成され、これらを決める必要がある。ここで、いかに質問を行うかは(1)と(2)のみに依存し、両者を合わせ質問戦略と呼ぶ。(3)の仮説の決定方法を定めたもとでは、学習効率と学習精度は質問戦略に依存し、どの質問戦略を用いるかが重要である。

これまで、質問からのブール関数の学習問題は、学習可能性を厳密に解析する計算論的学習理論の立場から多くの研究がなされている。これらの研究では学習の精度基準と効率基準を定めたもとで、様々な種類の質問方法と様々なブール関数クラスに対し、どの組合せであれば両基準を満たす学習が可能かどうかを明らかにしている。しかし、実際に学習を行う場合、有限の質問回数に対し、何らかの意味で最適な質問戦略を用いることが好ましいが、ブール関数の学習に関する従来研究では両基準を満たす質問戦略が存在するかどうかを明らかにすれば十分であり、最適な質問戦略を求めるといった研究はこれまで見られなかった。

本研究では、ベイズ決定理論により質問からの学習問題を組合せ最適化問題として定式化することで最適な質問戦略を定義し、最適な質問戦略を求めるアルゴ

リズムを与えることを目的とする．ベイズ決定理論は機械学習の分野で近年用いられているが，その理由は有限長のデータ系列に対して何らかの意味で最適な決定を行いたいという目的に対し，理論的にも実務的にも良く適合するためである．なお，ベイズ決定理論が用いられているのは機械学習の中でも例からの学習であり，質問からのブール関数の学習にベイズ決定理論を用いた研究はこれまでなされていない．本研究は，質問からの学習にベイズ決定理論を用いることで有限時点での最適性を議論可能な枠組みを導入し，組合せ最適化問題としての新たな定式化を与えている．

以下に本論文の構成を示す．

第 1 章では，本研究の序論として，研究の背景・目的について述べる．

第 2 章では，本研究で必要となる問題設定について説明する．まず，機械学習と質問からの学習のモデルについて説明し，次に従来研究の課題について述べる．ここでは特に最適化から見た従来研究と本研究の違いを明確にする．

第 3 章では，ベイズ決定理論により質問からのブール関数の学習問題を組合せ最適化問題として定式化を行う．学習の目的は真のブール関数を推定することとし，効率基準に質問回数，精度基準にベイズ決定誤り率を用いる．ここで，最適な仮説の決定方法は質問結果に無矛盾で事前確率が最大な仮説を選択することとなるため，この決定方法を用いる．次に，質問戦略を評価する基準が必要となるが，質問回数とベイズ決定誤り率の間にはトレードオフの関係が存在するため，一方を固定したもとの他方を最小化することが考えられる．そこで本研究では制約条件のもとで，目的関数を最小化する以下の 2 通りの評価方法を提案する．

評価方法 1・・・制約条件：質問回数を k 回以下，目的関数：ベイズ決定誤り率．

評価方法 2・・・制約条件：質問結果が得られたもとの決定誤り率が ε 以下，目的関数：平均質問回数．

この 2 通りの評価方法を用い，質問からの学習問題は組合せ最適化問題として定式化されるが，実際に学習を行うには最適解である最適な質問戦略を求めるアルゴリズムが必要となる．以後，第 4 章では一般の事前確率分布に対するアルゴリズムを与え，第 5 章では機械学習で仮定される，ある特定の事前確率分布に対するアルゴリズムを与える．

第 4 章では，一般の事前確率分布に対する質問からの学習問題を扱う．組合せ最適化問題は，一般には探索により最適解を求めなければならない，困難な問題であることが知られている．分枝限定法は，規模があまり大きくない場合に対して最も用いられ，最適解を効率よく探索するアルゴリズムである．ここで分枝限定法を用いるには，目的関数の下界値が必要となるが，下界値がタイトである程効率がよい．本章では，評価方法 1，2 の両者に対し，それぞれタイトな下界値を提案している．このうち，評価方法 2 では，目的関数が平均質問回数であるため，

平均質問回数の下界値が必要となる．数学モデルで考えた場合，質問学習の平均質問回数と情報理論の平均符号長は等価であり，平均符号長のタイトな下界値はエントロピーで与えられることが **Shannon** により示されている．そこで，平均質問回数においてもタイトな下界値がエントロピーを用いた計算式で与えられることを定理として与える．そして，この定理により得られた下界値を用いた分枝限定法により，最適な質問戦略を効率よく求めることが可能であることを示す．

第 5 章では，評価方法 1 に対し，ある特定の事前確率分布に対する質問からの学習問題を扱う．第 4 章では分枝限定法により質問戦略を求めたが，一般に分枝限定法では規模の大きい問題に対しては適用することができない．しかし，ある特定の事前確率分布に対しては，ある特殊な入力集合を質問することが最適な質問戦略となる．ここで，特殊な入力集合とは直交計画と呼ばれる入力集合であり，線形独立な列ベクトル集合族を持つ行列により構成される．これより，ある特定の事前確率分布に対し最適な質問戦略を求める問題は，行列が満たす条件として線形独立な列ベクトル集合族が与えられたもとで，行数が最小となる行列を構成する問題に帰着することができる．行列を構成するための探索は必要であるが，第 4 章の探索に比べ容易であるため，より規模の大きい場合に対して適用可能であることを示す．また質問後，最適な仮説は一般には探索により求めなければならないが，直交計画を用いる場合は探索によらず計算により求めることができる．この計算には多次元の高速フーリエ変換を用いることができ，最適な仮説が効率よく求まることを示す．さらに，本章では本研究と符号理論の関係も明らかにしている．上記の行列は符号理論における t 重誤り訂正可能な線形符号や **UEP(Unequal Error Protection)** 符号の検査行列を一般化した行列となっている．この結果より，これまで行列の構成方法は有限射影幾何を用いた割付けが中心であったが，符号理論の結果を用いた代数的構造を取り入れた行列の構成方法も可能となる．

最後に，第 6 章において以上の成果をまとめ，結論と今後の課題を述べる．

研 究 業 績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
1. ○論文	直交計画を用いたブール関数の学習に関する一考察 電子情報通信学会論文誌(A) J86-A, No. 4, pp. 482-490, (2003-4) 浮田 善文, 松嶋 敏泰, 平澤 茂一
2. ○論文	質問からの学習問題の決定理論による定式化に関する一考察 情報処理学会論文誌, Vol. 39, No. 11, pp. 2937-2948, (1998-11) 浮田 善文, 松嶋 敏泰, 平澤 茂一
3. 論文	学習期間と制御期間に分割された強化学習問題における最適アルゴリズムの提案 情報処理学会論文誌, Vol. 39, No. 4, pp. 1116-1126, (1998-4) 前田 康成, 浮田 善文, 松嶋 敏泰, 平澤 茂一
4. 論文	矛盾を含む知識の取り扱いについての一考察 人工知能学会誌, Vol. 13, No. 2, pp. 252-262, (1998-3) 斉藤 幹也, 浮田 善文, 松嶋 敏泰, 平澤 茂一
5. ○論文	A Note on Learning of Probabilistic Hypotheses from Membership Queries Proceedings of International Symposium on Information Theory and its Applications, Vol. 2, pp. 604-607, Mexico City, MEXICO (1998-10) Y. Ukita, T. Matsushima and S. Hirasawa
6. ○論文	A Learning with Membership Queries to Minimize Prediction Error IEEE International Conference on Systems, Man, and Cybernetics, Vol. 5, pp. 4412-4417, Orlando, Florida, USA (1997-10) Y. Ukita, T. Matsushima and S. Hirasawa
7. 論文	A Note on Concept Learning with Membership Queries Proceedings of International Symposium on Information Theory and its Applications, pp. 308-311, Victoria, British Columbia, Canada, (1996-9) Y. Ukita, T. Matsushima and S. Hirasawa
8. 講演	直交計画と双対符号の関係に関する一考察 電子情報通信学会研究技術報告, IT2006-107, pp. 131-136, (2007-3) 浮田 善文, 松嶋 敏泰
9. 講演	質問学習に直交計画を用いた場合の予測アルゴリズムに関する一考察 人工知能学会研究会資料, SIG-FPAI-A402-12, pp. 25-32, (2004-11) 浮田 善文, 小泉 大城, 松嶋 敏泰, 平澤 茂一
10. 講演	質問学習と逐次実験計画の関係に関する一考察 電子情報通信学会研究技術報告, AI2003-63, pp. 1-6, (2004-1) 浮田 善文, 松嶋 敏泰, 平澤 茂一

研 究 業 績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
11. 講演	ブール関数の逐次実験計画を用いた学習に関する一考察 電子情報通信学会研究技術報告, COMP2002-52, pp. 47-53, (2002-11) 浮田 善文, 松嶋 敏泰, 平澤 茂一
12. 講演	ベイズ決定理論による定式化のもとで直交計画を用いたブール関数の学習に関する一考察 電子情報通信学会研究技術報告, COMP2001-59, pp. 51-58, (2001-11) 浮田 善文, 松嶋 敏泰, 平澤 茂一
13. 講演	フーリエ変換を用いたブール関数の学習に関する一考察 電子情報通信学会研究技術報告, COMP2000-56, pp. 49-55, (2000-11) 浮田 善文, 松嶋 敏泰, 平澤 茂一
14. 講演	質問からのブール関数の学習における学習戦略を求めるアルゴリズム 第 22 回情報理論とその応用シンポジウム予稿集, pp. 845-848, (1999-12) 浮田 善文, 松嶋 敏泰, 平澤 茂一
15. 講演	質問からの学習における予測誤りに関する一考察 第 20 回情報理論とその応用シンポジウム予稿集, pp. 697-700, (1997-12) 浮田 善文, 松嶋 敏泰
16. 講演	不確実性を持つ概念の質問による学習に関する一考察 人工知能学会全国大会（第 10 回）, pp. 75-78, (1996-6) 浮田 善文, 松嶋 敏泰, 平澤 茂一
17. 講演	質問を許す概念学習に関する一考察 第 18 回情報理論とその応用シンポジウム予稿集, pp. 537-540, (1995-10) 浮田 善文, 松嶋 敏泰, 平澤 茂一
18. 講演	PROLOG を対象とした帰納的学習の効率化 人工知能学会全国大会（第 8 回）, pp. 137-140, (1994-6) 浮田 善文, 松嶋 敏泰, 平澤 茂一
19. 講演	不確実な知識の演繹推論アルゴリズムに関する一考察 第 22 回情報理論とその応用シンポジウム予稿集, pp. 841-844, (1999-12) 小島 健司, 浮田 善文, 松嶋 敏泰
20. 講演	適合フィードバックによる情報検索に関する一考察 第 21 回情報理論とその応用シンポジウム予稿集, pp. 161-164, (1998-12) 鍛冶 美緒, 浮田 善文, 松嶋 敏泰
21. 講演	不確実な知識の推論における欠測データの取り扱いに関する一考察 第 21 回情報理論とその応用シンポジウム予稿集, pp. 153-156, (1998-12) 金澤 裕輔, 浮田 善文, 松嶋 敏泰

研 究 業 績

種 類 別	題名、 発表・発行掲載誌名、 発表・発行年月、 連名者（申請者含む）
22. 講演	音声認識における Hidden Markov Model のパラメータ推定に関する一考察 第 21 回情報理論とその応用シンポジウム予稿集, pp. 643-646, (1998-12) 高井 望, 浮田 善文, 松嶋 敏泰
23. 講演	マルコフ決定過程の計算アルゴリズムについて 人工知能学会全国大会（第 12 回）論文集, pp. 56-59, (1998-6) 井村 真樹, 浮田 善文, 松嶋 敏泰
24. 講演	対数線形モデルを用いた不確実な知識の推論法について 人工知能学会全国大会（第 12 回）論文集, pp. 308-311, (1998-6) 小島 健司, 浮田 善文, 松嶋 敏泰
25. 講演	The Generalization of Bayesian Network's Deductive Method 電子情報通信学会技術報告, IT98-32, pp. 25-30, (1998-7) D. Koizumi, Y. Ukita and T. Matsushima
26. 講演	直交表現された仮説の学習に関する一考察 第 20 回情報理論とその応用シンポジウム予稿集, pp. 701-704, (1997-12) 峯松 一彦, 浮田 善文, 松嶋 敏泰
27. 講演	決定を考慮したベクトル量子化法の提案 第 20 回情報理論とその応用シンポジウム予稿集, pp. 609-612, (1997-12) 高井 望, 浮田 善文, 松嶋 敏泰
28. 講演	未知パラメータを含むマルコフ決定過程における学習アルゴリズムの提案 第 19 回情報理論とその応用シンポジウム予稿集, pp. 597-600, (1996-12) 前田 康成, 浮田 善文, 松嶋 敏泰, 平澤 茂一
29. 講演	構成的帰納論理プログラムに関する一考察 電子情報通信学会技術研究報告, AI95-47, pp. 41-46, (1996-1) 新津 健, 浮田 善文, 松嶋 敏泰
30. その他	ブール環上での性質を用いた質問からの学習に関する研究 横浜商大論集, Vol. 37, No. 2, pp. 96-120, (2004-3) 浮田 善文
31. その他	質問からの概念学習に関する一考察 横浜商科大学紀要, Vol. 8, pp. 601-620, (2001-11) 浮田 善文